

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Dual-eye-tracking Vygotsky: A microgenetic account of a teaching/learning collaboration in an embodied-interaction technological tutorial for mathematics

### Permalink

<https://escholarship.org/uc/item/3d34w62z>

### Authors

Shvarts, A  
Abrahamson, D

### Publication Date

2019-09-01

### DOI

10.1016/j.lcsi.2019.05.003

Peer reviewed



# Dual-eye-tracking Vygotsky: A microgenetic account of a teaching/learning collaboration in an embodied-interaction technological tutorial for mathematics

Anna Shvarts<sup>a,\*,1</sup>, Dor Abrahamson<sup>b</sup>

<sup>a</sup> Lomonosov Moscow State University, Mokhovaya st., 11, bl. 9, Moscow, 125009, Russia

<sup>b</sup> University of California, Berkeley, Graduate School of Education, 2121 Berkeley Way, Office 4110, Berkeley, CA 94720-1670, United States of America

## ARTICLE INFO

### Keywords:

Attentional anchor  
Dual eye-tracking (DUET)  
Joint attention  
Mathematics education  
Micro-zone of proximal development  
Teaching/learning process

## ABSTRACT

Vygotsky conceptualized the teaching/learning process as inherently collaborative. We extend prior evaluations of this claim by enlisting eye-tracking instruments to monitor the perceptual activity of four teacher–student dyads, as the student solves a challenging manipulation problem designed to ground the scientific notion of parabolas in their new sensorimotor routines. Analyzing each dyad's gaze paths led us to model the teaching/learning process as the emergence and dynamic transformation of intersubjective coupling between the student and tutor perception–action systems. While the student's sensory-motor coordination gradually gravitates toward an effective routine, the tutor's perception is iteratively launched from the student's current action, until the tutor detects an optimal moment for verbal intervention. In this micro-zone of proximal development, the student's motor action comes to align with the tutor's cultural-perspective strategy. Our elaboration of the cultural–historical approach to teaching/learning draws on research on joint attention and joint action from the cognitive sciences as well as the embodied-design approach from the educational sciences and demonstrates a compatibility of Vygotsky's heritage and complex dynamic systems theory. Finally, we discuss the educational value of the observed student–tutor intersubjective coupling phenomena, thus grounding the contribution of this multidisciplinary study within educational concerns.

## 1. Overview

The objective of this study is to contribute to a better understanding of the teaching/learning process. Our approach is to evaluate and elaborate on Vygotsky's historical assertion that learning is a student and a teacher's irreducibly collaborative achievement, through which the student develops existing naive forms of perceiving and acting on the environment into cultural forms (Vygotsky, 1935/2001). One major challenge of this enduring effort to evaluate Vygotsky's model has been the lack of research instruments for making evident how the student and tutor are perceiving the world as they act upon it. Now, almost a century since Vygotsky first articulated his model, we are equipped with technological instruments and analytic methodology for tracking people's gaze direction at a high sampling frequency and searching for patterns in this sensory activity as they relate to co-performed physical actions. Recent

\* Corresponding author.

E-mail addresses: [a.y.shvarts@uu.nl](mailto:a.y.shvarts@uu.nl) (A. Shvarts), [dor@berkeley.edu](mailto:dor@berkeley.edu) (D. Abrahamson).

<sup>1</sup> Currently works at: Freudenthal Institute for Science and Mathematics Education, Utrecht University, PO Box 85.170, 3508 AD Utrecht, the Netherlands.

technological breakthroughs in simultaneously tracking *two* people's visual behaviors as they work on the same display (dual eye tracking, DUET) now enable us to investigate how a student comes to perceive the world and act on it similar to how the tutor perceives and acts on it.

Our evaluation of Vygotsky's assertion is situated in the empirical context of a technologically enabled educational activity designed for the discovery of mathematical notions related to parabolas. The student manipulates a virtual object in an attempt to solve a dynamical control problem, while the tutor offers some crucial guidance to signify the visual display in terms of its potential cultural meanings. The activity is an exemplar of the Mathematics Imagery Trainer activity genre (Abrahamson & Trninic, 2011). In this genre, students initially work on a non-symbolic, technologically enabled, embodied-interaction problem. Solving the problem fosters the emergences of new sensorimotor schemes that are believed to ground prospective mathematical concepts. Supported by the tutor's contingent guidance, the students subsequently adopt symbolic artifacts as frames of references for enhancing their solutions. In so doing, the students shift to mathematical semiotic expressions consonant with the learning objective and cultural meanings.

Our analyses will draw both on Vygotskian terminology (the zone of proximal development; real forms and ideal forms; cultural environment; student-teacher collaboration) and cognitive and educational science (joint action; joint attention; educational design; attentional anchor). In particular, we engage complex dynamic systems theory of cognition and development, by which we will model the tutor and student's collaborative achievement as coupling between their respective perception-action systems into a single distributed intersubjective system adapted to enact the social interaction task they jointly encounter. Indeed such a system corresponds to Vygotsky's claim that educational activity is an irreducibly collaborative process within a cultural environment. The student contributes to this process by enacting movements that enable the tutor to accumulate embodied experience of her learning and identify an auspicious moment for intervening, which we will be calling a *micro-zone of proximal development*. The tutor contributes by his active and sensitive perception of the student's progress, which enables him to usher the student to reframe her perceptual attention toward the focal objects.

In the literature analysis, we adopt recent findings from cognitive science that enhance our analytical capacity to interpret educational phenomena. We assert that our study contributes: (a) a theoretical elaboration of Vygotsky's conceptualization of the teaching/learning process as irreducibly collaborative—a conceptualization we endorse as the emergence and dynamic transformation of embodied intersubjective coupling between perception-action dynamic systems of a tutor and a student; (b) methodology for engineering, investigating, and visualizing this collaborative educational process; and (c) empirical exemplification of intersubjective coupling while a student and a tutor work on an interactive embodied task—the exemplification reveals prerequisites for the emergence and maintenance of an effective micro-zone of proximal development.

Section 2 presents a theoretical introduction to Vygotsky's view of the educational process and to relevant cognitive and educational science constructs. Section 3 will explain our methods, including the embodied-interaction task. The section will end with our research questions for the empirical part of the research. Section 4 then reports on our findings from analyzing behaviors of four tutor-student dyads who engaged in the task. Finally, Section 5 will offer our conclusions.

## 2. Theoretical introduction

Lev Vygotsky has made considerable contributions to educational theory through his post-dualist explication of the teaching/learning process. Whereas these contributions have been broadly disseminated and elaborated in the literature, there may still be room to revisit, deepen, and refine them in light of recent innovation in educational design and research methodology. We begin by outlining essential lines of Vygotskian thought on education. These lines of thought will then serve as thematic orientations to our empirical investigation that is engaging innovations in educational design and methodology.

### 2.1. Essential Vygotskian pedagogical ideas

Per Vygotsky, a child develops side by side with the adults who bear *ideal forms*, that is, culturally specific and contextual ways of perceiving reality and performing actions (Vygotsky, 1935/2001). These *ideal forms* prevail in children's social environment even as the children behave according to naïve, initial, or *real forms*; for children to be culturally competent, their real forms of behaving, perceiving, and understanding are to transform into ideal forms of cultural actions. The terms *real* and *ideal forms* must be understood here in a narrow sense as initial and final forms of learning, respectively. Transformation in children's development from real to ideal forms is thus unlike evolutionary process, such as phylogeny or anthropogeny, stresses Vygotsky, because it happens through the social co-enactment of the forms, that is, on occasions of collaborating with adults in co-performing the cultural practice in question.

Within Vygotsky's body of work, one can discern two lines of thought relevant to implementing his theory of learning and development in the form of educational research and practice. Both lines treat the child's personal constituting of meaning through engaging in cultural activity. Vygotsky's first line of thought is laid out in his early book "Educational Psychology." Therein, Vygotsky (1926/2001, 1997, p. 49) stresses that "the teacher educates the student by varying the environment." He continues to offer:

Just as a gardener would be acting foolishly if he were trying to affect the growth of a plant by directly tugging at its roots with his hands from underneath the plant, so is the teacher in contradiction with the essential nature of education if he bends all his efforts at directly influencing the student.

Vygotsky's proposal for transforming real forms into ideal forms posited a pedagogical constitution of ideal forms as latent in the deliberate organization of an environment according to specific learning objectives: subject matter content should be constituted such

that children would discover the knowledge by themselves.

In his later work, Vygotsky introduces a second line of thought to elaborate on his understanding of the intensive interaction between real and ideal forms within the child–adult communication: Children do not need to discover all cultural achievements on their own—ideal forms originally emerge in ontogenesis through “collaboration with others” (Vygotsky, 1935/2001, p. 90, AS translation) and later become individual functions. Importantly, this “collaboration,” which in both Russian and English literally means “collective labor,” implies dual-agent activity that is deeply vested in both a child’s and an adult’s participatory and interactive experience, including their sensory and physical experience. It is not the direct involvement of a child into a ready-made social practice; rather, it is considering the child as an active, constituting participant who negotiates and thus determines how the practice will develop no less than the adult does (Stetsenko, 2017).

In the remainder of this introduction, we dedicate sections to each of the following Vygotskian ideas concerning the teaching/learning process: the contradistinction of real and ideal forms (Section 2.2); the teacher as a creator of a pedagogical environment (Section 2.3); and educational activity as student–teacher collaboration (Section 2.4). For each idea, we will cite contemporary educational and cognitive-science research to elaborate Vygotsky’s understanding of the teaching/learning process. In particular, we will be highlighting studies of joint action and joint attention so as to propose a view of student–teacher collaborative activity as a close-knit interaction tending to establish an embodied coupling between perception–action systems.

## 2.2. Real (naïve) and ideal (cultural) forms of perception: Transformation through education

Education is generally viewed as changing a child’s knowledge and understanding. What may be less obvious is that education transforms the child’s perception. Vygotsky theorized the development of perception as transformation from real (or naïve) into ideal (or cultural) forms. In explicating this ontogenesis of perception, Vygotsky focused on the construct of meaning—whether it was absent or present in the child’s perception of a situation: “Experiments have shown that the meaningful character of the adult’s perception is not inherent in the perception of the child” (Vygotsky, 1987, p. 269). The adult’s perception is an “immediate fusion of the processes of concrete thinking and perception” (ibid.); as soon as speech develops and categorical thinking is added, “we can no longer separate the perception of the object as such from its meaning or sense” (ibid., p. 299). This development of psychological functions (i.e., thinking and perceiving) leads to a new unity, which Vygotsky calls “psychological systems.”

Ontogenetic transformation of perception from its naïve form has acquired diverse names in educational research literature: *professional vision* (Goodwin, 1994) stresses the context of “a domain of scrutiny” as well as the discursive practice in which perception emerges; *disciplined perception* (Stevens & Hall, 1988) adds the moral quality of being formed by others; *educated perception* (Goldstone, Landy, & Son, 2010) refers to the top-down transformation of perception when we learn to perceive something in accordance with higher-level cognitive needs; and *theoretical perception* (Radford, 2010) refers to the Marxist idea of social practice that transforms human senses.

Eye-tracking studies have repeatedly exposed this alleged difference between expert and naïve vision (for a review see Gegenfurtner, Lehtinen, & Säljö, 2011), the key difference being the ability of experts to discern relevant information more efficiently than novices (e.g., Haider & Frensch, 1996; Jarodzka, Scheiter, Gerjets, & van Gog, 2010; Krichevets, Shvarts, & Chumachenko, 2014). Case studies in mathematics education research reveal that perceptual expertise transforms the subjective construction and configuration of perceived objects. For example, when novices analyze linear graphs, they often pay attention to a line’s irrelevant micro-properties instead of its slope or its intersections with the axes (Arcavi, 2003). Children may not notice the spatial organization of patterns (Radford, 2010), or they may be confused in interpreting Cartesian coordinates, not realizing that counting begins at the zero point (Shvarts, 2018), and so on. Recorded eye movements confirm that “looking” does not mean “seeing” (cf. Stevens & Hall, 1988); this difference is also expressed by the terms “looking at” versus “looking through” in the context of technology enhanced learning, where the learner may not acquire the transferable meaning even while spatially following the guidance (Sharma, Alavi, Jermann, & Dillenbourg, 2017). Thus, per Vygotsky’s systemic model, the teaching/learning process must foster the child’s unification of meaning and sensuous image. Only thus will the ideal form come forth.

Experts can disclose for novices how they themselves attend to a “domain of scrutiny” (Goodwin, 1994) by highlighting its features critical to the inquiry or required actions (e.g., Jamet, 2014; Ozelik, Arslan-Ari, & Cagiltay, 2010), by animating these features (e.g., Boucheix, Lowe, Putri, & Groff, 2013), and through other forms of visual cueing (Abrahamson, Gutiérrez, Charoenying, Negrete, & Bumbacher, 2012). Yet students can learn from experts’ gaze behavior even in the absence of explicit didactical explanations. This educational phenomenon has been recorded in the domains of medicine (Gegenfurtner, Lehtinen, Jarodzka, & Säljö, 2017; Litchfield, Ball, Donovan, Manning, & Crawford, 2010), zoology (Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013), and in number-line estimation tasks (Gallagher-Mitchell, Simms, & Litchfield, 2017). In the more complicated domains of geometry and problem-solving (Hoogerheide, van Wermeskerken, Loyens, & van Gog, 2016; van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009), guidance by modeled gaze-paths was not effective. Perhaps the complexity of perceiving mathematical and logical diagrams requires the development of subjective perceptual activity (*how* to see) that cannot be accomplished only by external direct guidance (*where* to see).

By studying the ontogenesis of cultural perception, we hope to emulate Vygotsky in espousing his notion that educational research should “reveal to the teacher how developmental processes stimulated by the course of school learning are carried through inside the head of each individual child” (Vygotsky, 1987, p. 91).

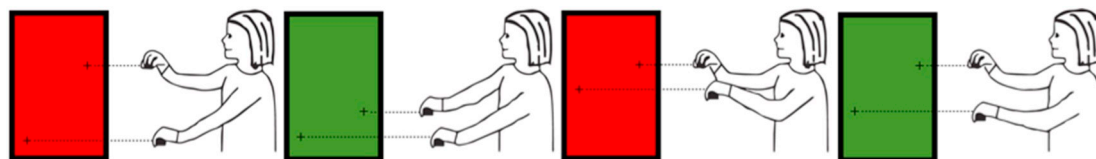


Fig. 1. The Mathematics Imagery Trainer for proportion. Here the screen is green when one cursor is two times higher above the bottom of the screen than another cursor, for a 1:2 ratio. Art acknowledgment: Virginia J. Flood. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 2.3. Creating a cultural environment for learning: Embodied educational design

Unlike pedagogical approaches that emphasize direct guidance of student attention within authentic arenas of professional practice, Vygotsky focused on the student's emergent forms of dynamical behaviors within dedicated cultural environments engineered for explorative activity. Per this pedagogical approach, disciplinary content should be introduced as a series of problems that lead a student to develop new and effective ways of acting in accordance with the task (Davydov, 1990).

The study reported in this article implements the pedagogical principle of teaching content as facilitating discovery. For this study, a learning environment was created in light of the embodiment turn in the theory of cognitive sciences (e.g., grounded cognition, Barsalou, 2003; enactivism, Varela, Thompson, & Rosch, 1991). In particular, it follows Abrahamson's (2014) action-based design genre, which specifies the development and facilitation of educational activities, in which students: (1) first learn to enact a new goal-oriented bimanual movement that instantiates a target mathematical concept; and only then (2) are guided to analyze this movement using mathematical instruments.

The Mathematics Imagery Trainer for proportions (Abrahamson & Trninic, 2011) is an environment that produces green feedback on a screen when a student's hands are placed on it at heights corresponding to an unknown ratio. Otherwise, the screen is red (Fig. 1). For example, when the technology is set at a 1:2 ratio, the screen will be green when the student's left and right hands are 10 and 20 cm above the bottom of the monitor, respectively. The student needs to learn to keep the screen green *while moving the hands*, which means that the hands must move at different speeds. Abrahamson and his collaborators have repeatedly observed students transition from the real form of interaction, in which the hands move at the same speed, to the ideal form, in which the hands move at different speeds. Slowly, in interaction with tutors and cultural artifacts, the students shift into ideal forms of discourse as well.

Students' perceptual strategies for manipulating the virtual objects according to task specifications are made manifest for the researcher through evidence of stable patterns of eye movements co-occurring with task-effective motor action (Duijzer, Shayan, Bakker, Van der Schaaf, & Abrahamson, 2017), which signify the emergence of sensorimotor coordinations; these coordinations are called *attentional anchors*. An attentional anchor is an imaginary perceptual construction that comes forth for individuals engaged in a demanding, ecologically coupled physical-control problem as their means of facilitating the coordination of sensorimotor schemes (Abrahamson & Sanchez-Garcia, 2016; Hutto & Sánchez-García, 2015). From the perspective of coordination dynamics (Kostrubiec, Zanone, Fuchs, & Kelso, 2012), the attentional anchor can be traced by state changes in an order parameter that marks a complex system's phase transition to a new dynamically stable constitution. The attentional anchor has been compared to Piaget's construct of reflective abstraction, in the sense that it constitutes a new psychological structure that coordinates and subsumes existing schemes to cope with a new class of situations (Abrahamson, Shayan, Bakker, & Van der Schaaf, 2016). In turn, the attentional anchor, as a subjective phenomenological entity, may bear what Bartolini Bussi and Mariotti (2008) call "semiotic potential," that is, it may come to constitute a mathematically meaningful system of relations. In the current project, the semiotic potential bears out through collaboration with a tutor: the emergence of an ideal form is facilitated by the tutor who encourages the students to reflect on their embodied activity and draw mathematical inferences (Morgan & Abrahamson, 2016). Rather than "directly suggest information" to the student, the tutor helps them "arrive at desirable insight" (Abrahamson et al., 2012, p. 78). As such, the learning process is explicated as the emergence of a new form of sensory-motor coordination, which later is conceptualized and sedimented by cultural semiotic means.

In this study, we corroborate Vygotsky's learning theory on the basis of empirical findings collected in the course of implementing an activity inspired by Abrahamson's design genre. We consider this educational design genre as suitable for operationalizing and evaluating Vygotsky's educational thesis and constructs, because the genre stresses the role of the student's personal activity and yet requires an indispensable collaboration with the tutor.

### 2.4. Learning as collaboration: Joint action and joint attention

Vygotsky viewed the teaching/learning process as principally different from natural development: The teaching/learning process, as a collaboration between an adult and a child, "creates a zone of proximal development" (Vygotsky, 1978, p. 90) and "awakens a variety of internal developmental processes that are able to operate only when the child is interacting with people in his environment" (ibid.); it is through this collaboration that new psychological functions, including new cultural meanings, evolve. Expanding Vygotsky's approach to education, Stetsenko writes: "Mind is always made in co-acting, together with other people, in shared collaborative activities that are part and parcel of wider social practices and collaborative projects" (Stetsenko, 2017, p. 319). The creative character of this collaboration might be illustrated by Vygotsky's example of the Russian author Tolstoy, as he teaches

creative writing to peasant children (Vygotsky, 2004): Tolstoy begins by writing a narrative exposition and then invites the children to think of a possible continuation to the narrative. Tolstoy acknowledges that children's abilities are so high, that *he* needs to learn from *their* natural inspiration. He adds that, "the children understood this real joint work with an adult writer to be a true collaborative effort, in which they felt themselves to be equal partners with the adults" (ibid., p. 49). This example stresses the principal feature of collaboration: an inherent co-action and mutual development that suggests we should consider a student and a teacher as forming an *interpersonal functional system* of ZPD (Newman, Griffin, & Cole, 1989), and rethink of ZPD as *symmetrical* (Roth & Radford, 2010).

Flood, Harrer, and Abrahamson (2016) have compellingly demonstrated that when a tutor and student engage in an embodied-interaction mathematics learning activity, together they reciprocally strive to maintain the shared intelligibility of the referents of their multimodal utterance, and as such they continuously mutually transform each other's perception of the domain of scrutiny. The role of this continuous mutual transformation is seen in the development of verbal labeling as well: the student and the tutor begin with vague references (prospective indexicals, Goodwin, 1996) to yet-unarticulated emerging functional entities that have not yet matured as bonafide mutually intelligible objects (Flood et al., 2016). Tutors' tactics, as they facilitate student engagements in embodied mathematics activities, were demonstrated as comprising repeating, revoicing, and elaborating on the student's multimodal utterance, both verbal and gestural (Abrahamson et al., 2012; Flood, 2018; Flood & Abrahamson, 2015). In each case, the tutor does not offer the student a ready-made explanation but instead adjusts his multimodal flow to the student's ad hoc expressions of her own experience. The contingency of this collaboration might explain why implementing the action-based design genre in artificial tutoring system is still challenging (Abdullah et al., 2017; Pardos, Hu, Meng, Neff, & Abrahamson, 2018). Alternatively, the tutor might intervene physically by initiating distributed co-performance of a task or even manually guiding the student's manipulations of the activity's implements (Abrahamson & Sanchez-Garcia, 2016). The authors characterize the tutor–student cooperation as a case of "intersubjective sensorimotor coordination by anticipating and closely tracking each other's actions" (ibid., p. 230).

The next section, below, cites contemporary literature on joint action and joint attention to propose the strong hypothetical claim that tutors' tactics invariably rely on intersubjective sensorimotor coordination, even when these tutor behaviors are not publicly evident. This paper provides a first step in empirical evaluation of this claim.

#### 2.4.1. Joint action: Two people coupled in one perception–action system

How do a tutor and a student engage in co-action, and what is the nature of any learning that transpires through this joint action? Cognitive research on *joint action* has been making strides. Numerous studies have demonstrated how people engaged in joint activity coordinate their actions (Knoblich, Butterfill, & Sebanz, 2011). Whereas some of these coordinations are planned, others unintentionally emerge during the joint action, such as the synchronicity of two people observing each other tapping (Oullier, de Guzman, Jantzen, Lagarde, & Scott Kelso, 2008) or rocking on rocking chairs (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007).

A dyad's capacity to co-act suggests at least partially overlapping co-orientation toward the action space, for example a shared imaginary number line that determines the speed of actions in the case of jointly sorting even and odd numbers (Atmaca, Sebanz, Prinz, & Knoblich, 2008). Jointly acting partners can also anticipate and endorse other properties of the co-actor's physical action, such as its destination, trajectory, and speed (Schmitz, Vesper, Sebanz, & Knoblich, 2017).

Whereas one form of joint action is mirroring (like joint tapping or synchronized reaching movements), another form is complementary actions, when each individual plays their own role. For example, two people who carry a table together must coordinate the performance of different constituent movements of the common action: One carries the table in front of themselves and the other carries it from behind (Sebanz, Bekkering, & Knoblich, 2006).

The most radical demonstration of asymmetrical joint action is provided by cases in which one person is required to follow the actions of someone else. Research shows that in these cases the observer anticipates the other's movements, preceding them with eye gaze (Flanagan & Johansson, 2003; Gredebäck & Falck-Ytter, 2015); these results might be attributed to simulative activity of the observer's mirror neurons system (e.g., Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Moreover, this other's-action anticipation relies on visual cues from different parts of the actors' bodies (Vaziri-Pashkam, Cormiea, & Nakayama, 2017), thus exemplifying the complexity of embodied interaction between two co-actors.

All these data support the idea of intersubjective coupling between perception–action systems (Spivey, Richardson, & Dale, 2009) as constituting an inter-individual dynamical system. Enactive functioning of this shared system might be analyzed within the dynamical-systems approach, with a focus either on: (a) self-organization (Dale, Fusaroli, Duran, & Richardson, 2014); or (b) the "extended body" (Froese & Fuchs, 2012), a neurophenomenological construct that offers an analytical means for characterizing the human capacity to synchronize behavior patterns with others by aligning rhythms and dynamics. Returning to the educational context, these theorizations, when applied to teaching/learning collaboration, appear to cohere with Vygotsky's approach and extend it.

Similar to the case of complementary co-action, such as carrying a table together, the two participants in a tutor–student dyad perform different parts of the action according to their different social statuses and knowledge, in so doing simultaneously enacting real versus ideal forms, respectively. Sharing the common action, teacher and student play different roles, where the student attempts to perform the required actions, while the teacher encourages, guides, or constrains them. Importantly, teaching/learning co-action aims not only to achieve some external pragmatic outcome of fulfilling some task demand but also to transform the learner's form of acting and perceiving. *Joint attention* (see below), considered an important mechanism in joint action (Böckler & Sebanz, 2013; Fiebig & Gallagher, 2013), might clarify how a student and a tutor come to perceive visual display in a similar way, including emerging mathematical objects.



#### 2.4.2. Joint attention: Guiding and following a student's attention

The construct of joint attention captures the phenomenon of two (or more) people focused on the same object and each aware of the other's focus. Over the past several decades, joint attention has been intensively investigated in research on early childhood and autistic disorders, and it has been found to be highly important for successful language acquisition (e.g., Tomasello & Farrar, 1986) and the development of social skills (e.g., Mundy & Newell, 2007) and other cognitive functions (e.g., Kopp & Lindenberger, 2011). It thus appears safe to conjecture that joint attention plays a key role in later learning as well: As soon as we analyze learning within a teaching/learning collaboration, the classic triadic relation structure between a learning object, a teacher, and a student inevitably includes joint attention.

Notwithstanding, joint attention has rarely been studied in educational contexts. Novel dual-eye-tracking (DUET) technology, which enables researchers to record the eye movements of two participants simultaneously, opens new horizons in the analysis of joint attention. Already several studies of joint attention during peer collaboration have used DUET (e.g., Belenky, Ringenberg, & Olsen, 2014; Schneider et al., 2016), including online learning with MOOCs (Sharma, Caballero, Verma, Jermann, & Dillenbourg, 2015). Interesting results were obtained in research involving joint play of Tetris (Jermann, Nussli, & Li, 2010), where pairs of experts and novices were investigated. It was shown that the perceptual actions of both the experts and the novices became more similar to each other as compared to the corresponding eye movements of the participants in expert-expert or novice-novice pairs. The teaching/learning process caused mutual transformation of the strategies.<sup>2</sup>

Investigations of initiating joint attention versus responding to joint attention in infants and adults in experimental laboratory settings show greater impact on learning outcomes when it is the child/learner who initiates joint attention than when the other does (Kim & Mundy, 2012; Tomasello & Farrar, 1986). One possible explanation is the motivational or affective factors that are inter-related with social leading and sharing: when a child/learner initiates and leads joint attention, positive affect increases (Grynspan, Martin, & Fossati, 2017; Hecke et al., 2007), and another explanation cites the enhanced feeling of agency that has been related to successfully initiating joint attention (Stephenson, Edwards, Howard, & Bayliss, 2018). Nevertheless, DUET analysis of infant-adult in vivo communication has revealed that joint attention can be initiated either by a child or the adult (Yu & Smith, 2016); and both patterns are present specifically in teaching/learning situations (Shvarts, 2018).

Traditionally, gaze following, namely an explicit inspection of the partner's *gaze direction*, was considered to be a key mechanism leading to joint attention (Bock, Dicke, & Thier, 2008; Brooks & Meltzoff, 2005). However, DUET revealed evidence of another important mechanism for establishing common focus, namely attention to the *manipulation of an object* (Yu & Smith, 2017; see also Pagnotta, 2018), thus coordinating between the gaze of one participant and the *actions* of the other participant. Analysis of different interactive paths leading to joint attention suggests a model of joint attention as resulting from self-organizing multimodal coupling between co-acting partners. The co-action is understood as an "exquisite real-time 'dance' of social interactions, in which effective adjustments within the dyad happen in fractions of seconds" (Yu & Smith, 2016, p. 3). In consonance with phenomenological and enactivist approaches, joint attention does not require "mind minding" (Hutto, 2011). Rather, each participant relies on the immediate perception of the shared *living space* as this perception reveals the affordances for action and tunes into the intentions and dispositions of the other participants' actions directly apparent in their gestures and movements (Gallagher, 2011; see also Meyer & Von Wedelstaedt, 2017). One of the key mechanisms of such direct perception of the environment is *strong anticipation*. The theoretical construct of strong anticipation is based on rejecting historical models of adaptive biological activity as constituted through the organism's ongoing complicated calculations of future events and, instead, proposing that the activity unfolds through tight enactive coordination with relevant features of the environment (Steepe & Turvey, 2010), including others' actions.

Strong anticipation may emerge in teaching/learning collaborative dyadic activity through deep alignment between a student's intentionality to solve a task and the tutor's enculturated "know-how" (*intentional synthesis*,<sup>3</sup> see Shvarts & Zagorianakos, 2016). This alleged importance of tracing the other's intentionality is supported by experimental data evidencing that by observing someone's movement, humans are able to anticipate the arc of their intentionalities even from early kinematic information (Becchio, Manera, Sartori, Cavallo, & Castiello, 2012; Pesquita, Chapman, & Enns, 2016). Moreover, as humans observe somebody's actions, they do not follow it but predict the next move (Flanagan & Johansson, 2003; Gredebäck & Falck-Ytter, 2015). Thus a directly visible partner's point of contact with a situation in question and prediction of their next move might turn out to be a simple, ecologically relevant cue for obtaining joint visual attention.

This theoretical section began by revisiting Vygotsky's view of cultural knowledge and practice, and in particular high psychological functions, as constituted by individuals through collaborative purposeful action with more knowledgeable others. We then cited contemporary data from educational and cognitive science to cast the teacher/student achievement of purposeful collaborative action as contingent on establishing an intersubjectively coupled perception-action system. This coupling emerges as a mechanism of joint attention and joint action, as they are fulfilled by dynamical embodied coordination between the tutor's and student's attention and actions. This process of interweaving two individual systems into an intersubjective system may serve the function of occasioning contingent reactions by the tutor to the student's actions and utterances, thus building the teaching/learning process with mutual respect to the tutor's cultural guidance and the student's personal activity.

We now shift to the empirical sections of our paper, where we offer evidence in support of our re-view of Vygotsky's educational thesis. In particular, we will enlist clinical and DUET data gathered from implementing an action-based educational design to argue

<sup>2</sup> For a computer-based simulation of experts adjusting to novices' performance characteristics, see Abrahamson and Wilensky (2005a, 2005b). Run the simulation here: <https://tinyurl.com/Piaget-Vygotsky>

<sup>3</sup> The intentional synthesis is drawn from the Husserlian phenomenological perspective on mathematics education (Zagorianakos, 2015).

for the existence and pedagogical value of embodied intersubjective coupling in tutor–student interaction. Our analyses of the participants' perceptual actions will reveal how joint attention is covertly attained, sustained, and yet manipulated in the course of dyadic co-action toward new learning achievements.

### 2.5. Research questions

We are proposing a neoVygotskian model of pedagogical interaction as the intersubjective dynamical coupling between perception–action systems. As we turn to evaluate this model empirically, we formulate the following general questions as a means of orienting our research project toward organizing its methods: What kind of evidence for intersubjective coupling between perception–action systems might be traced in multimodal data that include dual eye-tracking, video, and audio records of a student and a tutor participating in an embodied action-based mathematical activity? How does this coupling contribute to the teaching/learning process? In operationalizing the construct of intersubjective coupling between perception–action dynamical systems, we are particularly focused on evidence that might come from:

1. Similarity and dissimilarity between student and tutor eye-movements, as these would operationalize real and ideal forms of perceptual strategies;
2. The emergence of synchrony and co-localisation between student and tutor eye-movements, as these would operationalize joint attention—its genesis and sustainment;
3. Coordination between student action and tutor perception, as this would operationalize irreducible coupling between student and tutor systems.

## 3. Methodology

### 3.1. Dual eye-tracking technology

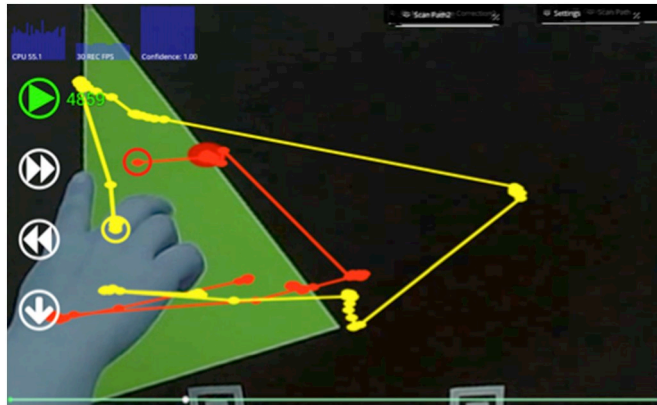
Dual eye-tracking (DUET) is an emerging instrument that initially appeared circa 2005 in the fields of cognitive linguistics (Richardson, Dale, & Kirkham, 2007), collaborative problem solving (Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008; Pietinen, Bednarik, Glotova, Tenhunen, & Tukiainen, 2008), and collaborative learning (Sangin, Molinari, Nüssli, & Dillenbourg, 2008). DUET allows synchronous tracking of the eye movements of two people seated in front of the same monitor (e.g., Pietinen et al., 2008; Shvarts, 2018), in front of two synchronized monitors (e.g., Jermann et al., 2010; Richardson, Marsh, et al., 2007; Sharma et al., 2015), or acted in a flexible environment (e.g., Macdonald & Tatler, 2018; Pfeiffer & Renner, 2014; Schneider et al., 2018). Whereas DUET holds potential for the analysis of ongoing teaching/learning collaborations by capturing participants' synchronous attention to features of the visual scene, multimodal analysis of joint performance in the shared space is still rather challenging (for a discussion, see Shvarts, Stepanov, & Chumachenko, 2018).

In our research, we use two Pupil-Labs (Kassner, Patera, & Bulling, 2014) eye-tracking goggles, which are head-mounted eye-trackers. Head-mounted eye-trackers allow freedom of movement in ecological settings where two people share a common space and discuss manipulations, such as on a shared monitor (Fig. 2). Thus, DUET supplements on traditional audio and video recording of two people engaged in an activity by synchronously capturing each participant's perception process as it is disclosed through capturing the directions of their gazes. The novelty of our DUET equipment (compare with Lilienthal & Schindler, 2017; Schneider et al., 2016)



**Fig. 2.** A dual eye-tracking experimental setting, where two participants discuss an image on a shared screen (a still from Stage 2 of the experimental activity).





**Fig. 3.** A sample of a video frame with an eye-tracking data. Each node-like point on a gaze-path line represents one sample, and the duration between two points is about 17 ms. The circles represent current eye-positions at the moment of this video frame.

is that we are able to overlay the scan paths of both participants' eye movements onto the video of their actions in the shared visual scene, including their manipulations and gestures (Shvarts et al., 2018). The sample rates of the eye-trackers (60 Hz, or roughly every 17 ms) and the world scene camera (30 Hz) allow us to grasp fairly well most of the events of interpersonal communication. Data records of the two participants' eye-tracking records are synchronized at about 1 ms.

This technical solution makes qualitative frame-by-frame analysis possible and efficient. Fig. 3 demonstrates one frame from the data. It shows a black screen with a green triangle on it and a gesture in front of the screen. Two lines on the frame represent the raw gaze paths of a tutor (red line) and a student (yellow line). Two circles (red and yellow) represent the positions of the gazes at the moment of this video frame. The gaze paths show the route of the gaze during the last two seconds, with each dot on the line representing the position of the gaze at a frequency of 60 Hz (in many cases the dots are close to each other, so they are compressed into a node, which corresponds to a fixation). By following the line, we can reconstruct the gaze trajectory: most recent positions are closest to the circle icon, while features that are most distant from the circle icon represent the position of the gaze 2 s prior the current frame.

### 3.2. Procedure and participants

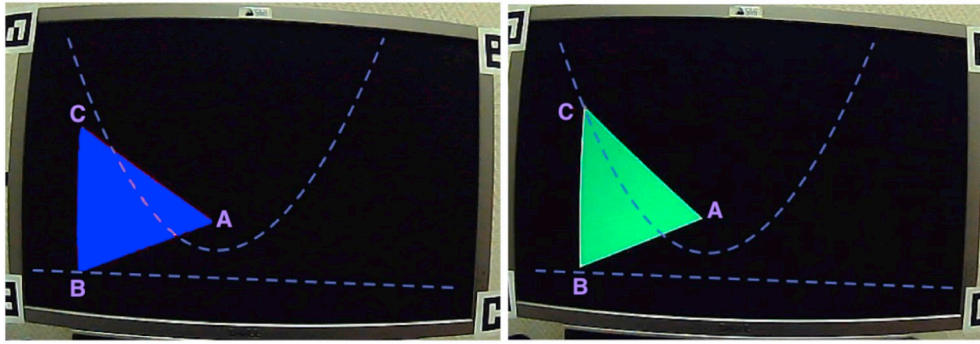
Four 17–21 year old undergraduate psychology students took part in our dual eye-tracking study. All the students had completed secondary school at a middle level of mathematics proficiency, however they could be considered as novices in relation to the learning task: as it was confirmed at the end of the study, they had no immediate memories of the parabola equation and had never studied a formal definition of a parabola as a locus of points equidistant from a line and a point. All the students were invited to participate in an eye-tracking study of mathematics teaching and learning.

There were two pairs, and each of those went through the following procedure: The tutor and the student wore eye-tracking goggles and took turns going through a pre-task 5-point calibration procedure. Then the task was run, and the student was free to manipulate the triangle on the screen by finger movements on a touchpad laying on the desk. The three task stages (see below) were run consecutively. The first student performed a parabola activity, with the researcher (the first author) serving as tutor. Next, the first student became a tutor for a second student, who had not been present during the first stage. The phases that we state in the Results and Discussion section (see below) were observed in all sessions, whether the researcher or an instructed student served as tutor.

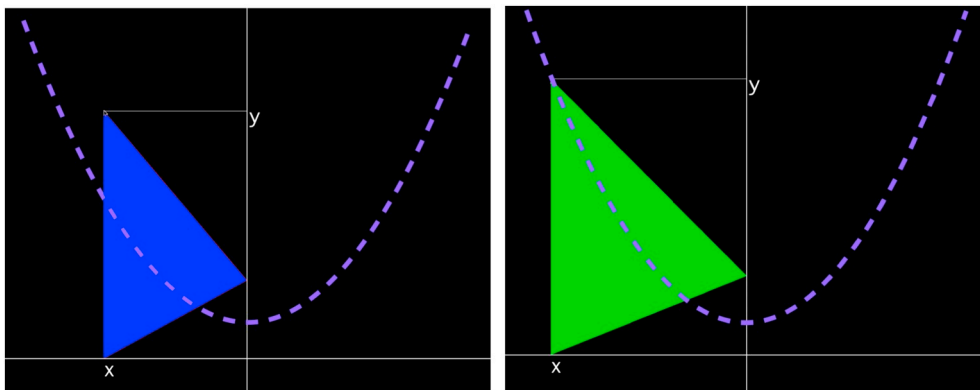
### 3.3. Learning activity

An interactive activity was constructed following principles of the action-based embodied design genre (Abrahamson, 2014). In this activity genre, the student interacts physically with a responsive technological platform. The student is tasked to discover on the screen a set of loci that satisfy a goal state, such as causing an object to take on a particular property. In this study, students manipulated the vertex of a triangle (Point C in Fig. 4) in an attempt to make the triangle green. The collection of points that satisfy this task objective are located along a parabola. This solution results from the geometrical properties of parabolas that have been programmed into the activity as task constraints bearing particular numerical values: To make the triangle green, Point C should be positioned on the screen such that it is equidistant from a straight horizontal line (see Point B directly below Point C, running along the parabola's directrix) and a separate point (see Point A fixed on the parabola's focus); in this case the triangle ABC is isosceles. Consistent with the design principle that semiotic symbols should be absent in the initial embodied activity (Abrahamson, 2014), only a triangle is featured on the screen during the first stage of the task (the dashed lines in Fig. 4 are for illustration only and are never shown to the students). Our activity follows earlier work that has used the topic of parabolas as a productive resource for generating interesting empirical data related to the development of mathematical conceptualization from an embodied activity (Brown, Heywood, Solomon, & Zagorianakos, 2013; Zagorianakos & Shvarts, 2015).

The successful implementation of action-based embodied designs require collaboration with a tutor (Abdullah et al., 2017;



**Fig. 4.** An action-based embodied design for parabolas. The triangle is green when it is isosceles with  $BC = AC$ , where B runs along the horizontal dashed line, A is the parabola's focus, and the student manipulates Vertex C. By keeping the triangle green while moving Vertex C, the student would effectively be inscribing a parabola. Note that the labels (A, B, C) as well as the dashed lines in this figure are used only here to illustrate the design for readers of this text—these lines do not appear for the students as they engaged in the activity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 5.** The second stage of action-based embodied design for parabolas. Axes and markers for coordinates are introduced. (The dashed line was not exposed to the participants.) In this second stage, mathematical symbolization is introduced, including the appearance of the orthogonal axes of the Cartesian plane and X and Y markers for projections of Vertex C on the axes (Fig. 5). The students were guided by the tutors to determine the formula of the curve traced by the manipulated Vertex C in keeping the triangle green: they manipulated and discussed the images on the screen, and finally the student was required to write down the formula of the curve (Fig. 2). The third stage was dedicated to further practicing the use of the parabola formula, and in particular to coordinating between, on the one hand, symbolic elements of the parabola formula and, on the other hand, spatial relations among geometrical elements in the Cartesian display. For this purpose, a grid was added onto the monitor plane, and participants were requested to predict grid node locations of Vertex C for which the triangle would become green. This paper focuses only on the first stage of the activity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Abrahamson et al., 2012; Flood, 2018; Flood & Abrahamson, 2015; Flood et al., 2016; Pardos et al., 2018). The tutors in our study were instructed to help students without giving them explicit solutions.

In the first stage of the activity, the tutor asked the student to “make the triangle green” (otherwise, it is blue); once they succeeded, the tutor suggested to “move the triangle in a way that it will always stay green.” The students are initially unaware of the rule that determines the triangle’s color—determining the rule for keeping the triangle green while manipulating Vertex C becomes their activity goal. The teaching/learning process transpires as a joint activity, eventually leading to the student discovering the rule through active exploration.

### 3.4. Data analysis

We draw on Vygotsky’s idea of theory–method dialectics, and we constitute our methodology within the triplet of theoretical principles, methods, and research questions (Radford & Sabena, 2015). So doing, we identify our unit of analysis as the *intersubjective* system of a tutor and a student engaged in a teaching–learning activity rather than their respective *individual* teaching and learning processes (see also Newman et al., 1989, pp. 59–75). Thus, we are specifically focused on intersubjective coordination and discoordination between: (a) student attention and tutor attention; and (b) student action and tutor attention. For an episode to be coded as manifesting joint visual attention, we judged the spatial and temporal alignment of eye-movements in the context of their channels of multimodal interaction (such as gestures and verbal utterances) to confirm that this alignment was an essential outcome of joint activity rather than occasional overlap of independent gazes. Spatial distance between the tutor’s and student’s gazes in joint

attention episodes could vary, taking into account the apparent object of reference: if the participants are discussing a triangle's particular side, they would need to gaze at this specific side so as to be coded as jointly attended, but if they are discussing the triangle as a whole entity, they may gaze at its different sides and still be coded as jointly attending.

We consider this analysis as a type of micro-ethnography (Streeck & Mehus, 2005), as we were in search of patterns across student actions, student/tutor gaze parameters, and the dyad's multimodal utterance. That said, the results presented in this paper—in Sections 4.1, 4.2, and 4.3—are the extracts of sensory-motor behavior that might be considered as preceding, or very early phases of, co-operative actions (Goodwin, 2017) or participatory sense-making (De Jaegher & Di Paolo, 2007). That is, no interaction is yet apparent as such in these data: the students act, while the tutors observe but do not intervene in any interpretative manner—they only rarely prompt further motor action, but without any topical commentary. Our analysis focused on documenting and modeling any coordination between the student's and the tutor's perceptions and the student's actions that became apparent through dual eye-tracking lens. So doing, we hoped to show how such coordination enables future interaction on joint sense-making (see Section 4.4.).

Unlike in the dialogues during the final phase of the interaction (Section 4.4.), the behavior of the participants in these earlier phases is mostly limited to visual and motor modalities: a student manipulates an object on the screen, while a tutor monitors the action. Whereas these episodes may appear as though “not much is happening,” in fact they bear cogent information for analyzing the formation of intersubjective coupling. Thus, we investigate the establishment of forms of intersubjective coordination that facilitate future interaction and co-operation.

#### 4. Results and discussion

Analyzing the video, audio, and eye-movements data from all four tutor–student pairs (of which 2 were researcher–participant pairs) led to the following 4-phases characterization of the activity within the first stage of the task.

##### 4.1. Phase 1: Convergence: The tutor tunes in to the student's real form

A student begins exploring the problem space for “green locations,” and so the triangle is blue most of the time and only occasionally becomes green, whence the student searches for further green positions. The student moves the triangle around the screen, while the tutor spontaneously follows it. Student and tutor gaze paths are tightly synchronized, as we explicate further.

Fig. 6 (see also Video 6) presents evidence of the student/tutor gaze synchronization. Recall that one person (the student) is performing the screen action, while the other person (the tutor) is only observing. The tutor's gaze lags after the student's gaze by about 17 ms, thus following precisely the student's movements of the triangle on the screen. In other cases this lag is between 30 and 60 ms, which is still briefer than the 80–150 ms for saccadic reaction times for a new visual stimulus, according to cognitive science experiments (Bompas, Hedge, & Sumner, 2017; Kirchner & Thorpe, 2006). These results correspond to findings that people tacitly predict the movements of others while observing their actions (Flanagan & Johansson, 2003; Gredebäck & Falck-Ytter, 2015).

At a later moment, a student has found a green location (Figs. 7 and 8). Her gaze turns ahead of the current position, apparently as she tries to predict the whereabouts of yet another “green location.”

Unlike Fig. 6, in Fig. 7 the tutor and student's gazes are not tightly synchronized. Instead, while the tutor's gaze is quite stationary, the student's gaze path suggests she is planning the next movements: her gaze goes ahead of her movement to the top-left of the screen, stays there for about 600 ms (Fig. 7a, b), then returns, focalizing on the triangle that she then moves along the trajectory foreshadowed by her gaze path (Fig. 7c). Eye-tracking allows us here to trace thinking-in-action. Vygotsky says, “Thought is not expressed but completed in the word” (1987, p. 250). But the word is not the only way to *act* as a person is thinking: we observe thinking as it is performed (not expressed) in eye-movements (cf. Grant & Spivey, 2003 for the essential role of eye-movements in insight problem-solving; see Kirsh & Maglio, 1994, on the role of physical rotations in the game Tetris).

The student continues to move the triangle (Fig. 8a, b), and about 200 ms later the tutor's gaze catches up with this movement (Fig. 8c). Thus, at unpredictable turns the tutor's gaze is delayed, taking 200–300 ms to catch up.

The data from Phase 1 have exemplified episodes of dynamical coordination within joint attention, as the tutor observes the students' early attempts to manipulate the triangle. The tutor's attention was either closely synchronized with the student's attention

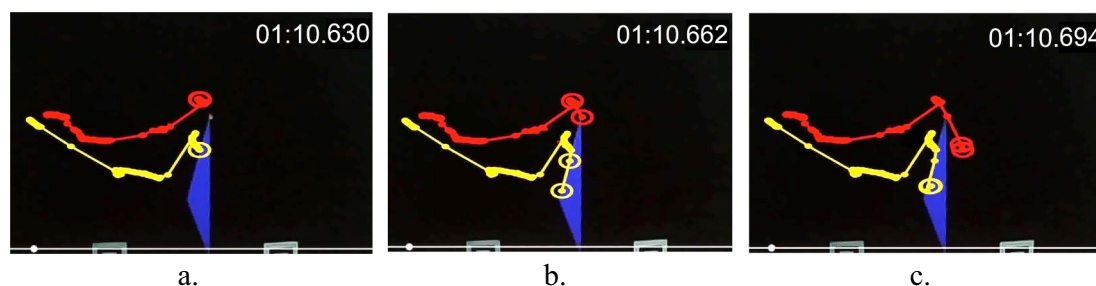
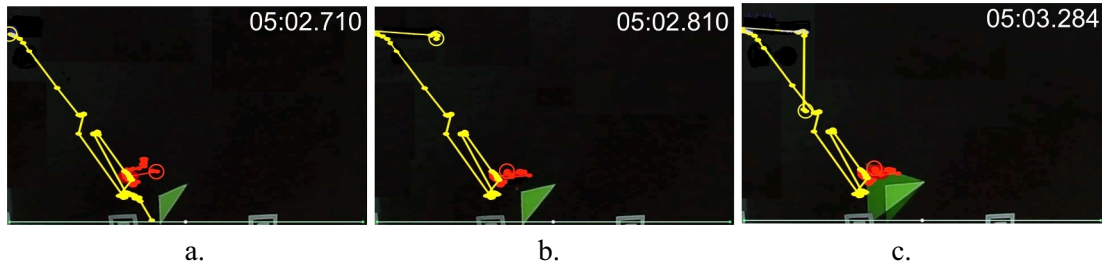
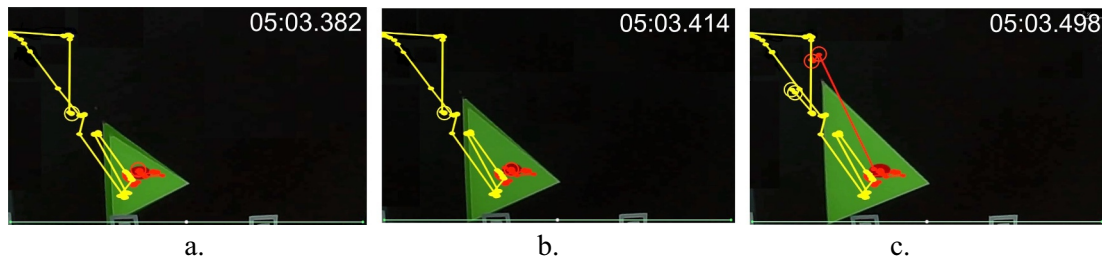


Fig. 6. Synchronous eye movements of the student (yellow) and the tutor (red), as the student manipulates Vertex C (the blue triangle's top vertex). Figs. 6a, b, and c show the sequential development of the gaze path. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 7.** The triangle is green. The student (yellow) plans the next movement, as her gaze rises to the left (a, b). She then moves the triangle to the same direction (c). The gaze of the tutor (red) remains stationary. Frame C presents the beginnings of the triangle move. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 8.** The tutor (red) catches up with the student's action: the triangle continues its movement up to the left (a, b, c), with the tutor's gaze following it at a 200 ms lag (c). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

or lagged slightly behind it, until an opportunity emerged to join the student's action and sustain the joint attention. As we have just seen, the tutor follows and anticipates the movements of the triangle on the screen. In so doing, the tutor appears to join tightly the student in her perception and action, while the triangle can be theorized as an extension of the student's body.

How do two people come to attend to the same objects at the same moment? Literature on joint attention has posited that gaze coordination is attained by one partner following the gaze of the other partner (Brooks & Meltzoff, 2014). However, we wish to eschew the potentially problematic epistemological underpinnings of an explanatory model that assumes one person second-guessing another person's reasoning structure (see critiques of "mind-minding" in Gallagher, 2011; Hutto, 2011). Instead, our data suggest that a tutor is following *the objects* operated by a student and anticipating *the objects'* state and/or location, as they are manipulated by a student. The speed of the events suggests that these "following" and "anticipating" are not deliberate moves but the results of the tutor's direct coupling with the environment, which includes the student's actions. This coupling between a student and a tutor is possible, we assume, on the basis of *strong anticipation* (Stepp & Turvey, 2010) and *intentional synthesis* (Shvarts & Zagorianakos, 2016). The tutor's and the student's perception-action systems are coordinated within the same environment, as they both share the embodied task and the possible ways to enact solutions. An attentive reader might complain that the tutor merely follows the triangle's movements and has nothing to do with the student's performance. In response, we would refer the reader to the phase 2, where the tutor demonstrates a perceptual strategy that goes far beyond mere observing the triangle.

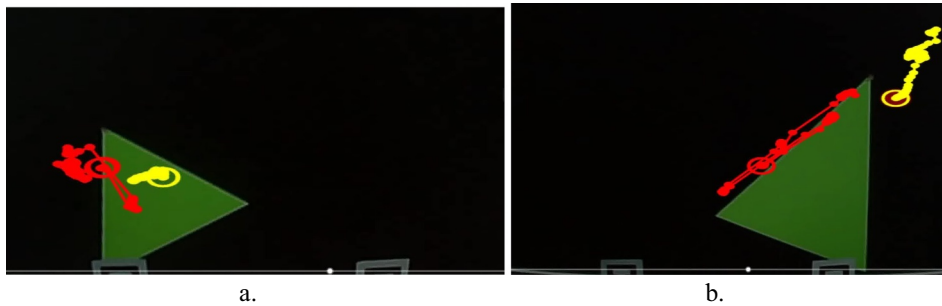
From our perspective, joint attention emerges naturally from joint action (see also Yu & Smith, 2017), without any would-be construction of the other's cognitive processes. In our case the student is controlling the manipulated object and thus leading the joint action and, respectively, the joint attention, while the tutor tunes in. According to some researchers (Kim & Mundy, 2012; Tomasello & Farrar, 1986), this student-led behavior may engender better learning outcomes in some tasks as compared to teacher-led behavior.

The consistent evidence in our empirical data of coordination between the *movement* (motor action) of one partner and the *gaze* (perception) of another allows us to consider these two people as two perception-action systems that are coupled intersubjectively into a single distributed intersubjective system, thus revealing the interaction as a dynamical self-organized system (Dale et al., 2014; Spivey et al., 2009). When broken by unpredictable movements, the coordination is rapidly restored and sustained. The tutor thus appears to experience the student's solution attempts vicariously. Importantly, this joint behavior partially corresponds to the Vygotskian idea of collaboration within the zone of proximal development, where the student is the main actor while the tutor is sensitive and open to the student's strategies and ready to adjust their own perception accordingly. However, at this stage we do not see a tutor contributing to the student's enculturation. Phase 4 (see far below) will demonstrate how these episodes of student-led joint attention and the tutor's embodied experience of the student's learning process play an important role in guiding the student to the cultural meaning.

#### 4.2. Phase 2: Divergence: The tutor tunes away from the real and into the ideal

Phase 2 begins once the students have found some green locations and have partially succeeded in keeping the triangle green





**Fig. 9.** The tutor performs perceptual patterns manifest in rapid, iterated, back-and-forth saccadic eye movements, while the student's eye movements are initially restricted to continuously tracing only the focal object they are manipulating (Vertex C of the moving triangle). Figures (a) and (b) provide the examples from different tutors. The red (tutor's) gaze path represents the forward and backward saccades over one iteration. The yellow (student's) gaze represent smooth pursuit behavior. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

while moving it but are not yet fluent in these movements, have not yet constituted a new sensorimotor coordination, and have not yet formulated their sensorimotor strategy in terms of a mathematical rule for keeping the triangle isosceles.

Unlike Phase 1, Phase 2 is not characterized by intersubjective gaze synchronization. On the contrary. While the students' gaze mostly continues to follow the operational point (Vertex C), the tutors change their perceptual strategy to a cultural form as soon as the triangle is green: they construe the green triangles as isosceles, thus manifesting an *ideal form* of perception. The tutors, who are not physically manipulating the triangle, thus activate the attentional anchors that they would employ *if they themselves were manipulating the triangle* from its current location that in fact is entirely under the student's command.

Fig. 9 exemplifies the tutors' attentional anchor as deployed either along the triangle's median (see also Video 9a) or along its side. While the tutors were not actually performing any action, their perceptive systems were coordinated with the actions of the students, consistently "launching" from the constantly shifting location of the triangle under the student's command, as though the tutor is practicing how they would perceive the triangle if they were manipulating it themselves.

The Phase 2 data set expands in several ways on the existing literature on attentional anchors (AA) in the learning of mathematics. First, while earlier AA have been demonstrated only in designs for proportions (Abrahamson & Sanchez-Garcia, 2016; Duijzer et al., 2017; Shayan, Abrahamson, Bakker, Duijzer, & van der Schaaf, 2017), we have presented evidence for AA in a new mathematical content domain, namely for parabolas. Second, we demonstrated "inert" AA, where individuals' (the tutors') perceptual behaviors were not consummated as bonafide actions. Apparently, their sensorimotor coordinations were elicited by observing the student's goal-oriented actions, a phenomenon which might be explained by the mirror neurons hypothesis (e.g., Rizzolatti et al., 1996).

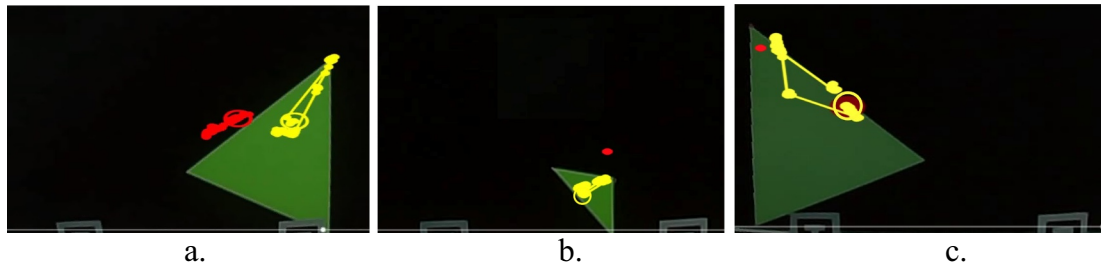
Importantly, though, the tutors' eye movements did not duplicate the student's eye movements as they had done beforehand, in Phase 1, or as demonstrated in earlier observations of simple movements (Flanagan & Johansson, 2003; Gredebäck & Falck-Ytter, 2015) that do not bear additional mathematical meaning. The green triangle held different meanings for the tutor and the student: The tutor's perception was coupled with the movement of the triangle perceived as an *action of keeping the triangle isosceles*, while the student's perception was coupled with the same movement but perceived as an *action of keeping the triangle green*, as the student did not yet intend for isosceles property. At Phase 3, as reported below, we will see that students finally do achieve an AA that is similar or even identical to the tutors'. However we remain with the question as to the specific pedagogical significance of Phase 2: What could possibly be the pedagogical rationale or intuition for this tutor-initiated temporal lapse in perceptual coordination?

During Phase 2, the tutor's *ideal form* of perception co-exists with the student's *real form* of perception, so that the perception processes are out of sync. Note, however, that two perception-action systems are still coordinated: the tutor's perception kept launching from the student's operational point, iteratively and rapidly updating to visual elements in the dynamic display that the student was handling. Thus, despite discoordination in *their perceptions*, the tutor's AAs reveal *coordination between the tutor's perception and the student's actions*, thus again supporting a model of student and tutor as dynamically coupled in a single distributed intersubjective system. We speculate that this asynchrony of eye movements enables the tutor to keep track of the student's activity development vis-a-vis the target performance that matches the tutor's ideal form. The student's performance needs to mature through their own activity within the cultural environment and educational design toward some critical performance criterion that would mark for the tutor an opportunity for productive intervention. Only then would the tutor's intervention target the student in a micro-zone of proximal development, thus leading to a truly collaborative discovery of the new meaning. The next two final Phases exemplify the transformation from the students' real form of perception toward the ideal form of perception, that is, the educated, professional, theoretical, disciplined perception (Goldstone et al., 2010; Goodwin, 1994; Radford, 2010; Stevens & Hall, 1988).

#### 4.3. Phase 3: Embodied discovery of the 'ideal': The student establishes a new sensorimotor coordination within the educational environment

How did the ideal AAs emerge in the student's activity and what was the tutor's role in this? Phase 3 is marked by the tutors encouraging the students by expressing appreciation of their movements and inviting them to move faster, slower, or more accurately. The video data manifest students' increasing motor fluency in manipulating the virtual objects, while the eye-tracking data





**Fig. 10.** Students' attentional anchors are reflected in rapid iterative saccadic eye movements either along the median of the triangle (a, b) or along one of its sides (c). (In the interest of this figure's clarity, the tutor's overlapping gazes have been removed from this image.)

reveal the systematic emergence of new AA facilitating the students' actions. Students' AAs are evidenced in their rapid, iterative saccadic eye movements either along the median of the triangle (see the yellow activity in Fig. 10a, b) or along one of its sides (Fig. 10c), concurrent with enhanced motor performance (greater accuracy and speed).

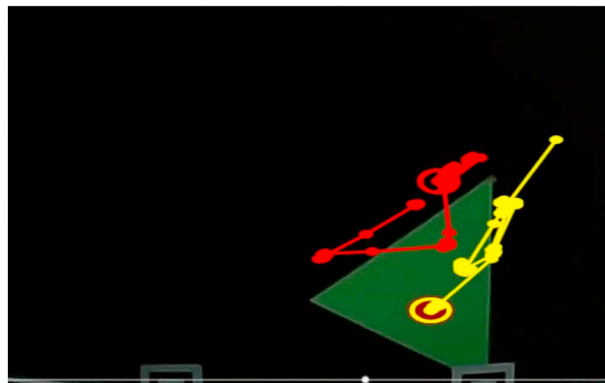
In Phases 1–3, the tutors' tactics were limited to supporting and encouraging the students' motor activity—students' development of the ideal form of mathematical perception improved not through receiving direct instruction but through enactment within the interactive design, with its embedded constraints and feedback. Notably, the activity task as stated is about action, not perception. It does not specify how students should construe the working environment, only that they achieve minimal fluency in manipulating the virtual objects within the task constraints (dynamically keeping the triangle green). Indeed, at no point do tutors indicate or suggest to the students specific productive ways of seeing the objects. As such, it could well occur that students achieve manual fluency with the task, even as they attend to the visual display different from their tutors who are sitting right by them and looking at their actions.

Our empirical data support the conjecture that the student and tutor's perceptual strategies develop independently, with evidence that the student and tutor were performing different iterative eye movements while watching the same triangle, thus engaging different AAs. Fig. 11 (see also Video 11) shows a tutor performing a complicated eye movement pattern that scans first one side of the triangle and then the other, as if verifying whether it is true that the sides are of the same length. Simultaneously, the student, who is controlling the triangle, also performs repetitive eye movements, but these eye movements are along the median of the triangle.

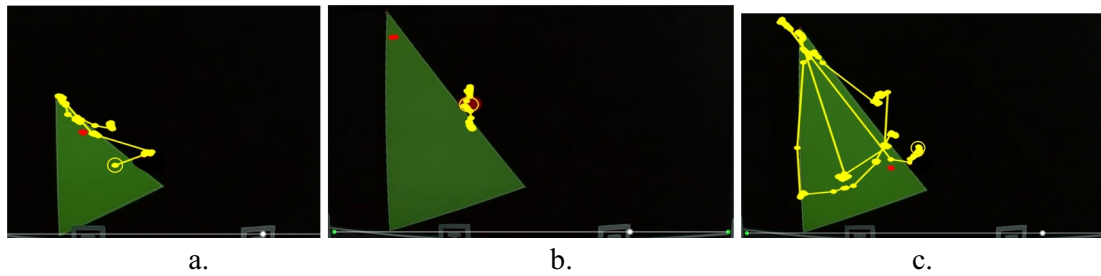
*Whereas the tutor and student need not construe the situation likewise, what needs to be coordinated is the mathematical meaning of their respective eye-movement strategies.* In the next section we will trace the emergence of consensual meanings during Phase 4.

#### 4.4. Phase 4: Collaborating on a conceptual discovery of the ideal: Orienting the student to the emerging cultural meaning

Phase 4 begins where the student is manipulating the triangle quite fluently and both tutor and student are enacting eye-movement path patterns corresponding to their respective AA. However, the emergence of an AA does not necessarily imply an *awareness* of the rule. Rather, witnessing that the student's fluent motor performance accords with the task specifications (keeping the triangle green) marks for the tutor that the student is ready for a new conceptualization, that is, the tutor judges that the conceptualization is within the micro-zone of proximal development. We provide two examples to illustrate the particularity of this pivotal intervention moment: it is only when the student's sensory-motor experience is sufficient that this pivotal intervention move is successful.



**Fig. 11.** The tutor (red) and the student (yellow) perform iterative eye movements, but these eye movements are different, signifying different attentional anchors: While the student performs saccades along the median of triangle, the tutor performs a three-steps repetitive eye-movements: she traces one side of the triangle (on the left) and then goes to the other side and back to the manipulated vertex. (Note that the red tutor's gaze path should be interpreted as offset to the left of its actual position, due to constraints of instrument sensitivity). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 12.** The student focuses attention on the triangle: (a) She explores the right side of the triangle; (b) The student momentarily pauses her actions while attempting to recall the geometrical term for the class of triangles that her actions have (unwittingly) been generating (viz. isosceles triangles); (c) The student answers the tutor's question about the meaning of the term "isosceles triangle." (The teacher's gazes are deliberately not depicted here).

The tutor (T1) then invites the student (S1) to reflect, as the following transcription exemplifies.

T1: Could you think about the triangle? How do you manage to keep it green?

S1: Alright... The triangle is obviously... (She explores only the right side of the triangle in Fig. 12a) Oh, I am bad with geometrical terminology...

T1: It's alright, you can explain, I will help you with the word.

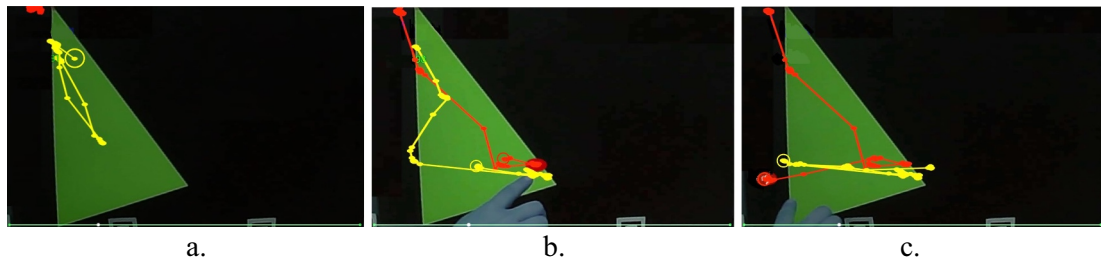
S1: It is not equilateral... but isosceles (Fig. 12b). I think that's how it's called.

T1: Yes, what does it mean?

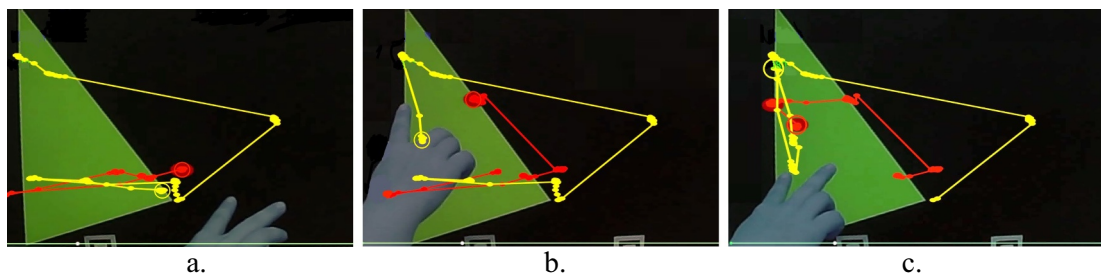
S1: It means... that it has two sides of equal length (Fig. 12c).

Interestingly, the student looked directly along the triangle's sides only *after* verbalizing the rule; she apparently already knew that the sides were equal and looked there perhaps to invoke or confirm the verbal explanation for the tutor.

Note that the tutor never directed the student toward any specific location within the triangle, a tutorial move that would have *redirected* the student's focus of attention. Instead, the tutor *reframed* the student's attention by referring the student to the triangle as a whole. *The tutor did not initiate a new joint attention episode but rather worked with the student's current focus of attention, reframing it.* The rule of keeping the triangle isosceles had already been surfaced, so that the student did not need to search for new features (Fig. 12 a, b).



**Fig. 13.** (a) The student's eye movements along the triangle's median before responding to the tutor's prompt; (b) and (c) The gesture and eye-movements that express the geometrical projection of one lateral side of the triangle onto the other.

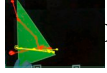


**Fig. 14.** The iterative gestures compare the sides as well as the gap between sides of an imaginary triangle on the right part of the screen (a) and the depicted triangle on the left part of the screen (b, c).

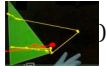
In the case of another student-tutor pair, the student (S2) did not arrive immediately at consensual meanings for her manipulation strategies. She responded to the tutor's prompt by mathematizing the AA (eye-movements along the median are seen in Fig. 13a, see also Videos 13–14) in a non-normative way:

T2: Can you think of a rule that keeps it green?

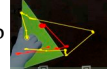
S2: I think it is a relation of this angle... [she gestures a projection from one lateral side to the other one (Fig. 13b), 13c



] Here, this angle will be the same [She compares the imaginary triangle on the right part of the screen (Fig. 14a



) with the triangle on the left part of the screen (Fig. 14b), 14c



the sides and the gap/angle between them].

T2: Which angle?

S2: I mean... the vertical..., the relation of the vertical side to the diagonal side, the one which is long.

T2: The angle is the same all the time?

S2: Seems... Yes...

T2: Even when it is like a line on the bottom [of the screen]?

S2: Yes... One moment... Let me check it! [The angle appeared to grow as the vertex was moved down.]

< ... >

T2: Try some more. Can you move it equable and smooth?

Thus, whereas this student achieved some task-effective sensorimotor coordination and we could trace an AA (Fig. 13 a), she did not verbalize this strategy correctly. Apparently her AA was established for moving the vertex only at the right and left parts of the screen: As we can see from the gestures (Fig. 14), she draws her conclusion from the comparison of lateral sides. The movements at the bottom part were not integrated into the same sensorimotor scheme. The tutor asks S2 to clarify her answer in detail and then, as the answer appears to be non-normative and rather confusing (the gestures and verbal utterances seems to have mixed references to the triangle's sides and angles), the tutor requires further motor practice.

Indeed, later the student overcame the confusion by crossing the bottom part of the parabola several times (with the tutor carefully following these movements). Later, the tutor *reframed* her attention precisely toward the bottom part of the curve, thus helping her discover the triangle's isosceles quality.

We have presented two examples of Phase 4, when the student has already achieved some task-effective motoric fluency and then the tutor suggests a reflection. That is, the interactive activity had transformed the student's action–perception coordination, a new attentional anchor has emerged, and now is the moment when the tutor's intervention could make a difference. We may theorize this moment as a *micro-zone of proximal development* (see also Kimmel, Irran, & Luger, 2014) that was created by the educational environment and the tutors' supporting activity. At this pedagogical moment, as in any zone of proximal development, collaboration greatly influences the student's abilities.

The first case demonstrated very gentle teaching guidance; it simply stops the student's activity for a moment and slightly reframes their attention without redirecting it. The tutor offers extra semiotic means (the verbal expression “Could you think about the triangle?”) that prompts for the student a new means of objectification (Radford, 2013); a new meaning for the visual scene emerges, and a new perceiving–thinking unity is formed (Vygotsky, 1987). This kind of teaching is possible, because the tutors closely followed the students' attention. The effectiveness of this kind of teaching is consistent with data on the pedagogical efficacy of responding to the attention of children (Tomasello & Farrar, 1986) or adults (Kim & Mundy, 2012).

The second case exemplified a moment of unsuccessful collaboration. The tutor prompts reflection before the student has integrated the move at the bottom of the curve with the right and left branches of the parabola and consequently the prompt fails to elicit the target response. The case demonstrates the tutor's lack of attention to the cumulative micro-history of the student's activity, which might have enabled the creation of a micro-zone of proximal development.

A juxtaposition of these two examples demonstrates the subtlety of the collaboration between student and tutor and why directly guiding the student's attention may have only limited effectiveness (e.g., Boucheix et al., 2013; Gegenfurtner et al., 2017; Hoogerheide et al., 2016; Jamet, 2014; Jarodzka et al., 2013; Litchfield et al., 2010; Ozcelik et al., 2010; van Gog et al., 2009). The ideal, meaningful form of mathematical perception emerges as a new, higher psychological function that systemically subsumes, adapts, integrates, and coordinates the earlier functions (Vygotsky, 1987). Namely the students' sensorimotor coordinations within the embodied educational environment, their earlier knowledge of the notion of isosceles triangle, and their inherent communicative function all made possible for the tutors to *reframe* the students' attention and to incorporate the cultural meaning into their embodied experience.

## 5. Concluding remarks and educational applications

Vygotsky revolutionized psychological and educational theory with a thesis that learning is the creative appropriation of cultural forms through social interaction that transforms personal activity. A century later, developments in philosophy, cognitive science,

pedagogical design, human–computer interaction technology, and scientific instrumentation have enabled us to visualize and monitor educational micro-processes accordingly as collaborative sensorimotor activity, wherein real and ideal forms tacitly tune to each other. Our data from action-based embodied mathematical activity offered an empirical operationalization of Vygotsky's constructs by comprising the respective simultaneous multimodal behaviors of a student and a tutor, including their perceptual actions (eye movements) and the student's motor actions (manipulating a virtual object on a screen) as well as clinical data of their interpersonal interaction from audio–video recordings.

The analysis, in accord with our neoVygotskian thesis, suggests a view of teaching/learning collaboration as embodied spatial–temporal coordination of two perception–action systems, wherein the tutor tunes into the real form of the student's action with high spatial and temporal synchrony (Phase 1). Later, in Phase 2, the student–tutor distributed coupled system undergoes major change, whereby the tutor departs from intersubjective perceptual synchrony, thus demonstrating co-existence of ideal and real forms within the same intersubjective activity. While the students are manipulating a screen object and either following its movements or scanning its prospective trajectory, the tutors' eyes begin performing a patterned sequence of eye-movements that launch from the object in motion, thus revealing their attentional anchors. This active mode of the tutor's observational process suggests their simultaneous cognizance of geometric structures on the screen and their anticipation of the student's current actions. The tutor thus gauges the quality of the student's actual performance as measured relative to the presumed quality of their own imaginarily projected actions. An educational environment shapes the student's action and perception during Phase 3 and determines the transformation of their perception strategies from the real form of following the triangle to the ideal form of emerging mathematically meaningful attentional anchors along with the fluency in motor practice. Tutors' involvement during this latter is limited to general support, as they wait for an optimal moment for intervention.

From a complex dynamic systems perspective, we theorize this irreducibly collaborative activity as the *coupling of two perception–action systems* into a single *distributed intersubjective system*. Joint attention is demonstrated as a mechanism that emerges from shared focus on the operational point of joint action. The revealed coupling of a tutor's perception with a student's actions is explained through the mechanisms of *strong anticipation*, namely a low-level ability of the tutor's perceptive system to distinguish relevant features of the student's performance. Change in relevant features from Phase 1 (just tracing a triangle) to Phase 2 (tracing a triangle as isosceles) demonstrates the importance of *intentional synthesis* beyond coupling, as the tutor's perceptive strategies are determined not by mere physical qualities of the display but by the cultural meanings of the student's actions.

Pedagogically, the coordination of a tutor's perception with a student's actions serves the need of sustaining the intersubjective coupling and co-constitution of the *micro-zone of proximal development* for further explicit tutorial intervention. The student–tutor embodied collaboration thus culminates in a highly deliberate “nudge,” wherein the tutor *re-frames* the student's attention so that the student elicits new meaning from their own praxis (Phase 4). Now, coordinated through verbally exchanged meanings, the tutor and the student may nevertheless tacitly differ in their subjective perspective strategies. This convergence at one level and divergence at another level corroborate the hypothetical systemic structure of acquired psychological functions, as it is drawn in Vygotsky's “psychological systems” claim as well as in the methodology of research on complex dynamic systems.

In sum, we have drawn on theoretical resources from cognitive science and educational research to investigate a body of dual-eye-tracking empirical data that could enable us to clarify and elaborate on Vygotsky's views on the micro-process of teaching–learning interaction. In appraising the educational implications of our study, we note the potential value specifically of attending to embodied dynamical mechanisms underlying and mobilizing educational processes such as scaffolding, student-oriented teaching, and the constitution of the zone of proximal development. As educational designers, we should be perspicacious in attending to these embodied teaching/learning processes, which include subtle, at times almost imperceptible, interactions between a student's and a tutor's sensory-motor actions. This awareness is especially important in the digital era, where a variety of novel technological solutions is touted to enhance the educational process. Considering the pivotal educational function of joint visual attention to an operational point of joint action, we may be more scrupulous in adjudicating the apparent potential of various technological innovation, such as multiple classroom screens or adaptive response to students' performance. More specifically, we might re-rationalize interface cues designed to redirect a student's attention toward particular screen loci of didactical import—perhaps rather than shifting the student's attention to a new locus, their attention should be reframed upon the current locus.

Further research is needed to clarify the generality of our findings: What conditions are conducive for the establishment of *intersubjective perception–action coupling*? In what other contexts might a tutor perform as *if* actions that fore/shadow a student's actions to detect an optimal moment for intervening in a micro-zone of proximal development? How often, and in which occasions, is *re-framing* a student's attention pedagogically effective?

This paper has contributed to educational research discourse on intersubjective perception–action coupling, interpreting the phenomenon through Vygotsky's legacy. Whereas the current state of evidence is idiosyncratic to the particular laboratory settings, we hope it opens directions for further investigation. As such, contemporary technological innovation, namely dual eye-tracking, being supplemented by multimodal data, provides an insightful lens for investigating student–tutor collaboration in the context of engaging in action-based embodied design for mathematics learning. Cognitive and educational science scholarship combined with the complex dynamical systems approach offer routes for honing Vygotskian literature on the teaching/learning process, while offering new horizons for educational practice informed by these theoretical innovations.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.lcsi.2019.05.003>.

## Acknowledgment

This work was supported by the Russian Foundation for Basic Research under Grant 18-013-00906.

## Declarations of interest

None.

## References

- Abdullah, A., Adil, M., Rosenbaum, L., Clemmons, M., Shah, M., Abrahamson, D., & Neff, M. (2017). Pedagogical agents to support embodied, discovery-based learning. In J. Beskow, C. Peters, G. Castellano, C. O'sullivan, I. Leite, & S. Kopp (Eds.). *Proceedings of 17th international conference on intelligent virtual agents (IVA 2017)* (pp. 1–14). Cham: Springer International Publishing.
- Abrahamson, D. (2014). Building educational activities for understanding: An elaboration on the embodied-design framework and its epistemic grounds. *International Journal of Child-Computer Interaction*, 2(1), 1–16.
- Abrahamson, D., Gutiérrez, J., Charoenying, T., Negrete, A., & Bumbacher, E. (2012). Fostering hooks and shifts: Tutorial tactics for guided mathematical discovery. *Technology, Knowledge and Learning*, 17(1–2), 61–86.
- Abrahamson, D., & Sanchez-Garcia, R. (2016). Learning is moving in new ways: The ecological dynamics of mathematics education. *The Journal of the Learning Sciences*, 25(2), 203–239.
- Abrahamson, D., Shayan, S., Bakker, A., & Van der Schaaf, M. F. (2016). Eye-tracking Piaget: Capturing the emergence of attentional anchors in the coordination of proportional motor action. *Human Development*, 58(4–5), 218–244.
- Abrahamson, D., & Trninic, D. (2011). Toward an embodied-interaction design framework for mathematical concepts. In P. Blikstein, & P. Marshall (Vol. Eds.), *Proceedings of the 10<sup>th</sup> annual interaction design and children conference (IDC 2011)*. Vol. "Full papers,". *Proceedings of the 10<sup>th</sup> annual interaction design and children conference (IDC 2011)* (pp. 1–10). Ann Arbor, MI: IDC.
- Abrahamson, D., & Wilensky, U. (2005a). Piaget? Vygotsky? I'm game!: Agent-based modeling for psychology research. *Paper presented at the annual meeting of the Jean Piaget society*. Retrieved February 8, 2018, from <http://ccl.northwestern.edu/research/conferences/JPS2005/jps2005.html>.
- Abrahamson, D., & Wilensky, U. (2005b). *NetLogo Piaget-Vygotsky Game model*. [tinyurl.com/Piaget-Vygotsky](http://tinyurl.com/Piaget-Vygotsky) Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University.
- Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational Studies in Mathematics*, 52(3), 215–241.
- Atmaca, S., Sebanz, N., Prinz, W., & Knoblich, G. (2008). Action co-representation: The joint SNARC effect. *Social Neuroscience*, 3(3–4), 410–420.
- Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358(1435), 1177–1187.
- Bartolini Bussi, M. G., & Mariotti, M. A. (2008). Semiotic mediation in the mathematics classroom: Artefacts and signs after a Vygotskian perspective. In L. D. English, M. G. Bartolini Bussi, G. A. Jones, R. Lesh, & D. Tirosh (Eds.). *Handbook of international research in mathematics education* (pp. 720–749). (2nd revised ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Becchio, C., Manera, V., Sartori, L., Cavallo, A., & Castiello, U. (2012). Grasping intentions: From thought experiments to empirical evidence. *Frontiers in Human Neuroscience*, 6(May), 1–6.
- Belenky, D., Ringenberg, M., & Olsen, J. (2014). Using dual eye-tracking to evaluate students' collaboration with an intelligent tutoring system for elementary-level fractions. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.). *Proceedings of 36<sup>th</sup> Annual Meeting of the Cognitive Science Society (CogSci 2014)* (pp. 176–181). (Quebec City, Canada).
- Bock, S. W., Dicke, P., & Thier, P. (2008). How precise is gaze following in humans? *Vision Research*, 48(7), 946–957.
- Böckler, A., & Sebanz, N. (2013). Linking joint attention and joint action. In J. Metcalfe, & H. Terrace (Eds.). *Agency and joint attention* (pp. 206–215). New York, NY: Oxford University Press.
- Bompas, A., Hedge, C., & Sumner, P. (2017). Speeded saccadic and manual visuo-motor decisions: Distinct processes but same principles. *Cognitive Psychology*, 94, 26–52.
- Boucheix, J.-M., Lowe, R. K., Putri, D. K., & Groff, J. (2013). Cueing animations: Dynamic signaling aids information extraction and comprehension. *Learning and Instruction*, 25, 71–84.
- Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106(3), 1465–1477.
- Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, 8(6), 535–543.
- Brooks, R., & Meltzoff, A. N. (2014). Gaze following: A mechanism for building social connections between infants and adults. In M. Mikulincer, & P. R. Shaver (Eds.). *Mechanisms of social connection: From brain to group* (pp. 167–184). Washington, DC, US: American Psychological Association.
- Brown, T., Heywood, D., Solomon, Y., & Zagorianakos, A. (2013). Experiencing the space we share: Rethinking subjectivity and objectivity. *ZDM Mathematics Education*, 45(4), 561–572.
- Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2014). The self-organization of human interaction. In B. H. Ross (Ed.). *The psychology of learning and motivation* (pp. 43–95). Elsevier Inc.: Academic Press.
- Davydov, V. V. (1990). In J. Teller (Ed.). *Types of generalization in instruction: Logical and psychological problems in the structuring of school curricula* Reston, VA: National Council of Teachers of Mathematics. Retrieved from <http://www.marxists.org/archive/davydov/generalization/generalization.pdf>.
- De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, 6(4), 485–507.
- Duijzer, C. A. C. G., Shayan, S., Bakker, A., Van der Schaaf, M. F., & Abrahamson, D. (2017). Touchscreen tablets: Coordinating action and perception for mathematical cognition. *Frontiers in Psychology*, 8, 144.
- Fiebig, A., & Gallagher, S. (2013). Joint attention in joint action. *Philosophical Psychology*, 26(4), 571–587.
- Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, 424(6950), 769–771.
- Flood, V. J. (2018). Multimodal revoicing as an interactional mechanism for connecting scientific and everyday concepts. *Human Development*, 61, 145–173.
- Flood, V. J., & Abrahamson, D. (2015). *Refining mathematical meanings through multimodal revoicing interactions: The case of "faster."* paper presented at the annual meeting of the American Educational Research Association.
- Flood, V. J., Harrer, B. W., & Abrahamson, D. (2016). The interactional work of configuring a mathematical object in a technology-enabled embodied learning environment. In C.-K. Looi, J. L. Polman, U. Cress, & P. Reimann (Vol. Eds.), *"Transforming learning, empowering learners," proceedings of the international conference of the learning sciences (ICLS 2016)*. Vol. 1. "Transforming learning, empowering learners," *proceedings of the international conference of the learning sciences (ICLS 2016)* (pp. 122–129). Singapore: International Society of the Learning Sciences.
- Froese, T., & Fuchs, T. (2012). The extended body: A case study in the neurophenomenology of social interaction. *Phenomenology and the Cognitive Sciences*, 11(2), 205–235.
- Gallagher, S. (2011). Interactive coordination in joint attention. In A. Seemann (Ed.). *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 293–305). United States: MIT Press.
- Gallagher-Mitchell, T., Simms, V., & Litchfield, D. (2017). Learning from where "eye" remotely look or point: Impact on number line estimation error in adults. *The Quarterly Journal of Experimental Psychology*, 218(August), 1–30.
- Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers & Education*, 113, 212–225.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523–552.
- van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human*



- Behavior*, 25(3), 785–791.
- Goldstone, R. L., Landy, D. H., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science*, 2(2), 265–284.
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96(3), 606–633.
- Goodwin, C. (1996). Transparent vision. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 370–404). Cambridge: Cambridge University Press.
- Goodwin, C. (2017). What is co-operative action, and why is it important? *Co-operative action* (pp. 1–20). Cambridge University Press.
- Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving. *Psychological Science*, 14(5), 462–466.
- Greddebäck, G., & Falck-Ytter, T. (2015). Eye movements during action observation. *Perspectives on Psychological Science*, 10(5), 591–598.
- Grynszpan, O., Martin, J. C., & Fossati, P. (2017). Gaze leading is associated with liking. *Acta Psychologica*, 173, 66–72.
- Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, 30(3), 304–337.
- Hecke, A. V. V., Mundy, P. C., Block, J. J., Delgado, E. F., Parlade, M. V., Meyer, J. A., & Neal, A. R. (2007). Infant joint attention, temperament, and social competence in preschool children. *Child Development*, 78(1), 53–69.
- Hoogerheide, V., van Wermeskerken, M., Loyens, S. M. M., & van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction*, 44, 22–30.
- Hutto, D. D. (2011). Elementary mind minding, enactivist-style. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 307–341). United States: MIT Press.
- Hutto, D. D., & Sánchez-García, R. (2015). Choking REctified: Embodied expertise beyond Dreyfus. *Phenomenology and the Cognitive Sciences*, 14(2), 309–331. <https://doi.org/10.1007/s11097-014-9380-0>.
- Jamet, E. (2014). An eye-tracking study of cueing effects in multimedia learning. *Computers in Human Behavior*, 32, 47–53.
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154.
- Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, 25, 62–70.
- Jermann, P. N., Nussli, M.-A., & Li, W. (2010). Using dual eye-tracking to unveil coordination and expertise in collaborative Tetris. In T. McWewen, & L. McKinnon (Eds.), *Proceedings of the 24<sup>th</sup> BCS interaction specialist group conference* (Dundee, United Kingdom).
- Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing adjunct publication (UbiComp '14)* (pp. 1151–1160). New York: ACM Press.
- Kim, K., & Mundy, P. (2012). Joint attention, social-cognition, and recognition memory in adults. *Frontiers in Human Neuroscience*, 6(172).
- Kimmel, M., Irran, C., & Luger, M. A. (2014). Bodywork as systemic and inter-enactive competence: Participatory process management in Feldenkrais® method and Zen shiatsu. *Frontiers in Psychology*, 5, 1424.
- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46(11), 1762–1776.
- Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4), 513–549.
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. In B. Ross (Vol. Ed.), *The psychology of learning and motivation*. Vol. 54. *The psychology of learning and motivation* (pp. 59–101). Burlington: Academic Press.
- Kopp, F., & Lindenberger, U. (2011). Effects of joint attention on long-term memory in 9-month-old infants: An event-related potentials study. *Developmental Science*, 14(4), 660–672.
- Kostrubiec, V., Zanone, P.-G., Fuchs, A., & Kelso, J. A. S. (2012). Beyond the blank slate: Routes to learning new coordination patterns depend on the intrinsic dynamics of the learner – experimental evidence and theoretical model. *Frontiers in Human Neuroscience*, 6. <https://doi.org/10.3389/fnhum.2012.00222>.
- Krichevets, A., Shvarts, A., & Chumachenko, D. (2014). Perceptual action of novices and experts in operating visual representations of a mathematical concept. *Psychology. Journal of Higher School of Economics*, 11(3), 55–78.
- Lilienthal, A. J., & Schindler, M. (2017). Conducting dual portable eye-tracking in mathematical creativity research. In B. Kaur, W. K. Ho, T. L. Toh, & B. H. Choy (Vol. Eds.), *Proceedings of the 41st conference of the International Group for the Psychology of mathematics education*. Vol. 1. *Proceedings of the 41st conference of the International Group for the Psychology of mathematics education* (pp. 231–). Singapore: PME.
- Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest X-ray inspection. *Journal of Experimental Psychology: Applied*, 16(3), 251–262.
- Macdonald, R. G., & Tatler, B. W. (2018). Gaze in a real-world social interaction: A dual eye-tracking study. *Quarterly Journal of Experimental Psychology*, 71(10), 2162–2173.
- Meyer, C., & Von Wedelstaedt, U. (Eds.). (2017). *Moving bodies in interaction – interacting bodies in motion: Intercorporeality, interkinaesthesia, and enaction in sports*. Amsterdam: John Benjamins.
- Morgan, P., & Abrahamson, D. (2016). Cultivating the ineffable: The role of contemplative practice in enactivist learning. *For the Learning of Mathematics*, 36(3), 31–37.
- Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, 16(5), 269–274.
- Newman, D., Griffin, P., & Cole, M. (1989). *The construction zone: Working for cognitive change in school*. New York, NY: Cambridge University Press.
- Oullier, O., de Guzman, G. C., Jantzen, K. J., Lagarde, J., & Scott Kelso, J. A. (2008). Social coordination dynamics: Measuring human bonding. *Social Neuroscience*, 3(2), 178–192.
- Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior*, 26(1), 110–117.
- Pagnotta, M. (2018). *Living and learning together: Integrating developmental systems theory, radical embodied cognitive science, and relational thinking in the study of social learning*. Doctoral dissertation University of St Andrews. Retrieved 10 November 2018 from <https://research-repository.st-andrews.ac.uk/handle/10023/16386>.
- Pardos, Z. A., Hu, C., Meng, P., Neff, M., & Abrahamson, D. (2018). Characterizing learner behavior from high frequency touchscreen data using recurrent neural networks. In D. Chin, & L. Chen (Eds.), *Adjunct proceedings of the 26th Conference on User Modeling, Adaptation and Personalization (UMAP '18)* (Singapore: ACM (6 pages)).
- Pesquita, A., Chapman, C. S., & Enns, J. T. (2016). Humans are sensitive to attention control when predicting others' actions. *Proceedings of the National Academy of Sciences*, 113(31), 8669–8674.
- Pfeiffer, T., & Renner, P. (2014). EyeSee3D: A low-cost approach for analyzing mobile 3D eye tracking data using computer vision and augmented reality technology. *Proceedings of the symposium on eye tracking research and applications - ETRA'14* (pp. 195–202). New York, New York, USA: ACM.
- Pietinen, S., Bednarik, R., Glotova, T., Tenhunen, V., & Tukiainen, M. (2008). A method to study visual attention aspects of collaboration. *Proceedings of the 2008 symposium on eye tracking research & applications - ETRA '08* (pp. 39). New York, New York, USA: ACM Press.
- Radford, L. (2010). The eye as a theoretician: Seeing structures in generalizing activities. *For the Learning of Mathematics*, 30(2), 2–7.
- Radford, L. (2013). Three key concepts of the theory of objectification: Knowledge, knowing, and learning. *Journal of Research in Mathematics Education*, 2(1), 7–44.
- Radford, L., & Sabena, C. (2015). The question of method in a vygotskian semiotic approach. In A. Bikner-Ahsbahs, C. Knipping, & N. C. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 157–182). Netherlands: Springer.
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science*, 18(5), 407–413.
- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26(6), 867–891.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Research. Cognitive Brain Research*, 3(2), 131–141.
- Roth, W.-M., & Radford, L. (2010). Re/thinking the zone of proximal development (symmetrically). *Mind, Culture, and Activity*, 17(4), 299–307.
- Sangin, M., Molinari, G., Nüssli, M., & Dillenbourg, P. (2008). How learners use awareness cues about their peer's knowledge? Insights from synchronized eye-tracking data. *8th international conference on international conference for the learning sciences (ICLS)*. Vol. 2. *8th international conference on international conference for the*

- learning sciences (ICLS) (pp. 287–294). International Society of the Learning Sciences.
- Schmitz, L., Vesper, C., Sebanz, N., & Knoblich, G. (2017). Co-representation of others' task constraints in joint action. *Journal of Experimental Psychology: Human Perception and Performance*, 43(8), 1480–1493.
- Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning*, 1–21.
- Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. D. (2016). Detecting collaborative dynamics using mobile eye-trackers. In C.-K. Looi, J. L. Polman, U. Cress, & P. Reimann (Vol. Eds.), "Transforming learning, empowering learners," *proceedings of the international conference of the learning sciences (ICLS 2016)*. Vol. 1. "Transforming learning, empowering learners," *proceedings of the international conference of the learning sciences (ICLS 2016)* (pp. 522–529). Singapore: International Society of the Learning Sciences.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70–76.
- Sharma, K., Alavi, H. S., Jermann, P., & Dillenbourg, P. (2017). Looking THROUGH versus looking AT: A strong concept in technology enhanced learning. In É. Lavoué, H. Drachler, K. Verbert, & J. Broisin (Eds.). *Data driven approaches in digital education. EC-TEL 2017. Lecture notes in computer science* (pp. 238–253). Cham: Springer.
- Sharma, K., Caballero, D., Verma, H., Jermann, P., & Dillenbourg, P. (2015). Looking AT versus looking THROUGH: A dual eye-tracking study in MOOC context. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Vol. Eds.), "Exploring the material conditions of learning: Opportunities and challenges for CSCL," *the proceedings of the computer supported collaborative learning (CSCL) conference*. Vol. 1. "Exploring the material conditions of learning: Opportunities and challenges for CSCL," *the proceedings of the computer supported collaborative learning (CSCL) conference* (pp. 260–267). Gothenburg, Sweden: ISLS.
- Shayan, S., Abrahamson, D., Bakker, A., Duijzer, C. A. C. G., & van der Schaaf, M. (2017). Eye-tracking the emergence of attentional anchors in a mathematics learning tablet activity. In C. Was, F. Sansosti, & B. Morris (Eds.). *Eye-tracking technology applications in educational research* (pp. 166–194). Hershey, PA: IGI Global.
- Shvarts, A. (2018). Joint attention in resolving the ambiguity of different presentations: A dual eye-tracking study of the teaching-learning process. In N. Presmeg, L. Radford, W.-M. Roth, & G. Kadunz (Eds.). *Signs of signification: Semiotics in mathematics education research* (pp. 73–102). Dordrecht: Springer.
- Shvarts, A., Stepanov, A., & Chumachenko, D. (2018). Automatic detection of gaze convergence in multimodal collaboration: A dual eye-tracking technology. *The Russian Journal of Cognitive Science*, 5(3), 4–17.
- Shvarts, A., & Zagorianakos, A. (2016). Theoretical perception of the Cartesian plane: A dual eye-tracking study through double theoretical lenses. In C. Csikos, A. Rausch, & J. Sztányi (Eds.). *Proceedings of 40th annual meeting of the International Group for the Psychology of mathematics education (PME 40)*, 3–7 august 2016 (pp. 287). (Szeged, Hungary).
- Spivey, M. J., Richardson, D. C., & Dale, R. (2009). The movement of eye and hand as a window into language and cognition. In E. Morsella, J. Bargh, & P. M. Gollwitzer (Vol. Eds.), *The psychology of action. Vol. 2. The psychology of action* (pp. 225–249). New York: Oxford University Press.
- Stephenson, L. J., Edwards, S. G., Howard, E. E., & Bayliss, A. P. (2018). Eyes that bind us: Gaze leading induces an implicit sense of agency. *Cognition*, 172, 124–133.
- Stepp, N., & Turvey, M. T. (2010). On strong anticipation. *Cognitive Systems Research*, 11(2), 148–164.
- Stetsenko, A. (2017). *The transformative mind*. Cambridge: Cambridge University Press.
- Stevens, R., & Hall, R. (1988). Disciplined perception: Learning to see in technoscience. In M. Lampert, & M. L. Blunk (Eds.). *Talking mathematics in school* (pp. 107–150). Cambridge: Cambridge University Press.
- Streeck, J., & Mehus, S. (2005). Microethnography: The study of practices. In K. Fitch, & R. Sanders (Eds.). *Handbook of language and social interaction* (pp. 381–404). Mahwah, NJ: LEA.
- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 57(6), 1454–1463.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: M.I.T. Press.
- Vaziri-Pashkam, M., Cormiea, S., & Nakayama, K. (2017). Predicting actions from subtle preparatory movements. *Cognition*, 168, 65–75.
- Vygotsky, L. S. 1926/2001. *Educational psychology* (R. H. Silverman, Trans.). Boca Raton, FL: CRC Press LLC.
- Vygotsky, L. S. (1935/2001). *Lektsii po pedagogii [Lectures on pedagogy]*. Izhevsk: Udmurtskii universitet (Original work partially published 1935).
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press (Original work published 1930).
- Vygotsky, L. S. (1987). Collected works of L. S. Vygotsky. In R. W. Rieber, & A. S. Carton (Vol. Eds.), *Problems of general psychology. Vol. 1*. New York: Plenum Press.
- Vygotsky, L. S. (2004). Imagination and creativity in childhood. *Journal of Russian and East European Psychology*, 42(1), 7–97.
- Yu, C., & Smith, L. B. (2016). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science*, 41, 1–27.
- Yu, C., & Smith, L. B. (2017). Hand-eye coordination predicts joint attention. *Child Development*, 88, 2060–2078.
- Zagorianakos, A. (2015). *The study of intuition as an objectifying act from a Husserlian perspective, in the cases of three prospective teachers of mathematics*. Doctoral dissertation Manchester Metropolitan University. Retrieved November 10, 2018 from [https://e-space.mmu.ac.uk/595477/1/Zagorianakos-PhD%20Thesis\\_final.pdf](https://e-space.mmu.ac.uk/595477/1/Zagorianakos-PhD%20Thesis_final.pdf).
- Zagorianakos, A., & Shvarts, A. (2015). The role of intuition in the process of objectification of mathematical phenomena from a Husserlian perspective: A case study. *Educational Studies in Mathematics*, 88(1), 137–157.